

HUMAN LIMITATIONS IN TOWED ARRAY SONAR RECOGNITION

S. M. McFadden*

M. M. Taylor*

Table of Contents

- 1. Introduction**
- 2. Capabilities of the Ear**
 - 2.1 Detection**
 - 2.2 Simple Discrimination**
 - 2.3 Binaural Processing**
 - 2.4 Change Rates of Amplitude and Frequency**
- 3. Processing of Sonar Sounds for Aural Display**
- 4. Integration of Information**
 - 4.1 Bisensory Displays**
 - 4.2 Automatic Processing**
 - 4.3 Cueing**
- 5. Proposed Aural Displays**
 - 5.1 Identification**
 - 5.2 Detection**
- 6. Conclusions**
- 7. References**

*Defence and Civil Institute of Environmental Medicine, P.O. Box 2000,
Downsview, Ontario, Canada, M3M 3B9

SUMMARY

Modern towed array sonar systems provide an enormous amount of information, both in the number of directions simultaneously scanned and in the information coming from any one direction. For the operator to be able to handle the incoming information, it must be presented in a concise and highly discriminable form: traditional high resolution visual displays take up considerable display space, and in a towed array system are suitable only for the analysis of stable narrow-band sounds. Lower resolution visual displays allow information from multiple sources to be presented, but again only for sources that are very stable. Time varying and broad-band signals are not well detected. The human ear is particularly suited for recognizing subtle changes in intensity and frequency and for integrating the energy from multi-component and broad-band sounds. The ear is best at dealing with signals whose energy is primarily in the range 300-3000 Hz, and whose variations occur over periods of 20-200 milliseconds. To match the capabilities of the ear, sea sounds must be transformed both in frequency and time. With the required speed-up, an operator would be able to listen to one suspicious sound repetitively, or to scan a large portion of the ocean during the time of a single update. The use of time-compressed aural displays in conjunction with current display modes should result in more efficient and more effective presentation of the information collected by towed-array systems.

1. INTRODUCTION

A towed array sonar system (TASS) allows a ship to detect targets at greater distances than did the previous generation of hull-mounted sonar systems, because it allows sources from different sectors of the ocean to be segregated. A target is masked only by sounds from the same apparent bearing, instead of competing with sounds from all bearings. Not only can quieter targets be detected than heretofore, but also more complex characteristics of the sounds from a given target can be discriminated.

To realise the benefits of the improvement in signal-to-noise ratio (SNR) afforded by the towed array, the operator should monitor the incoming signals from each separate direction (or beam). There may be 40 or more such beams in a typical TASS, each of which might show one or more important targets. In addition, the movements of known targets must be monitored, and the identities of newly detected ones ascertained. To aid the operator, the incoming information should be presented in a concise and highly discriminable form. The selection of processing techniques and displays is critical. The information must be compressed in some way so that the operator has a chance to evaluate it. Insufficient compression makes it difficult for the operator to handle all the information. Too much compression, or inappropriate compression, will lead to missed targets and the advantages of the towed array system will not be realised.

000001

In the early days of sonar, the operator just listened to what was going on in the sea. There were several disadvantages to this procedure:

- Noise on the ship interfered with the operator's ability to hear the interesting sounds;
- Much of the information was contained in steady tones, which are detected relatively poorly by the ear;
- The frequencies of interesting sounds were not necessarily those for which the ear is best suited.

Visual displays do not suffer from most of the above problems. Currently, they are the primary medium for presenting information to the operator. More recently, computer-driven detection and classification has been improving to a level at which it is useful to the operator. It is not yet, however, trusted to outperform a human operator in analysis of any given target; its strength is its ability to monitor all beams tirelessly and continuously.

The most effective form of visual display for single channel passive sonar systems has been a frequency-time-intensity (FTI) display. Each point on the display is the amplitude of the radiation over a narrow range of frequencies (frequently 1 Hz or less) averaged over a short period (one second or more) of time. Targets appear as a series of lines (usually vertical) against a noisy background. The lines represent the acoustic frequencies at which the energy received is greater than the background. Best detection is achieved when the frequency resolution of the system matches the bandwidth of the target lines, but the finer the resolution, the more processing time is required. Identification of targets is based on a knowledge of the pattern of frequencies emitted by different types of targets. Certain frequency ratios or the existence of certain highly stable frequencies are important clues to the identity of targets. These values are easily read off visual displays using mechanical or electronic aids if the relevant lines are visible.

In a TASS system, there is insufficient display space to permit optimum FTI displays for all beams at once, or even to scan all beams sequentially with normal resolution in frequency and duration in time. The high resolution FTI is suitable for classification only. For initial detection, preferred displays include automatic alert systems, bearing-time-intensity displays (BTI), and multiple FTI displays with limited frequency resolution and short histories. These displays compress the information so that the operator can scan all bearings quickly. Automatic alert systems compress data by allotting to a computer the task of examining all beams for suspicious combinations of frequencies. The automatic alert system and the multiple FTI display show the narrow-band stable sounds. The BTI, which shows the overall amplitude of the incoming radiation on each beam summed over frequency, presents some broad-band information. The operator scans the visual displays to confirm or reject the alerts of the automatic alert system and, it is hoped, to pick out signals that the automatic alert system has missed.

000002

As with aural displays, visual displays have their limitations, especially in a TASS system.

- FTI displays are best for the detection and classification of stable, narrow-band sounds. Not all sounds radiated by ships and submarines are stable narrow-band sounds. There are also transients and broad-band sources, which might be modulated over time. Broad-band sounds usually do not radiate sufficient energy at individual frequencies to be readily seen on an FTI display. Transient sounds are averaged out during signal analysis, and appear only as horizontal marks indistinguishable from normal sea noise. All these effects occur within a single beam of the TASS.
- Broad-band displays such as the BTI use fixed bandwidth filters. The result is frequently either the inclusion of unwanted noise, which degrades the SNR, or the exclusion of useful information at frequencies outside the filter boundaries.
- In order to display the output from multiple bearings in a limited display space, the information is compressed over time and frequency. The reduced resolution means a poorer SNR, which reduces detection capability.

To enhance the capability of the operator using TASS, the discriminability of targets to be classified must be increased. Increased discriminability leads to faster and more accurate classification, and simplifies tracking. To improve detection, display resolution should be improved and reliance on purely visual presentations reduced.

Auditory displays can present information not readily available on visual displays. At the same time, they reduce the load imposed on the visual system. Increased resolution is achieved without an increase in display space. However, as mentioned above, aural displays have their own limitations, and should never be considered as substitutes for visual displays. As with visual displays, preprocessing is needed to achieve an effective aural display.

This paper examines the possibility of developing aural displays that could be used with existing visual displays and automatic alert systems to make more effective use of the TASS operator. The capabilities of the auditory system are examined with a view to determining the best ways of transforming the signals from the TASS. Methods of enhancing incoming sounds are suggested that would improve their discriminability and apparent information content. In addition, studies on the integration of information from different senses are reviewed to determine how aural displays might best be used in concert with automatic alerting systems and visual displays.

2. CAPABILITIES OF THE EAR

This section reviews relevant psychoacoustical studies in detection and discrimination of simple sounds. These studies mostly provide upper bounds on the capabilities of the auditory system, since they were performed in laboratories under controlled conditions. Under operational conditions, the signals are more complex, and the interfering sounds are different from those of the laboratory. Nevertheless, these fundamental studies provide data on the basic capabilities of the aural system. An effective aural display exploits what the ear does best.

2.1. Detection

The first and most important question about a sound is "Can it be detected?" Studies in detection measure the ability of the ear to "hear out" a sound (usually called the signal) in the presence of other sounds (usually called maskers or noise). The detectability of a given signal in a given masker can be predicted by comparing the effective energy in the signal with the effective energy in the noise (signal to noise ratio or SNR). The term "effective" is used because not all the energy in the signal or the noise may be used by the listener. The effective energy in a given signal depends on the spectrum of the signal and the masker, the duration of the signal, and the listener's knowledge of both signal and background.

In detecting a signal masked by noise, the ear acts as if it were listening only to a narrow band of frequencies around the frequency (or frequencies) of the signal. This band of frequencies is called the "critical band". The critical band has been defined in different ways, but all definitions come down to two versions with associated experimental approaches: The first is that the critical band is the band of frequencies that affects the detectability of a signal (Version 1), and the second is that it is the band of frequencies over which the ear can integrate signals perfectly (Version 2). Studies on version 1 change the bandwidth of a masking noise and estimate the critical band to be that bandwidth beyond which increases in masker bandwidth do not affect detectability of a tonal signal in the centre of the band. Such studies usually find that the "critical band" is about a sixth of an octave [1]. Studies on version 2 concentrate on the signal rather than the masker, and determine the bandwidth within which tones interact either in detection or in pitch estimation [2,3]. There are fewer such studies, but they tend to find smaller values for the critical band, around a twelfth of an octave.

The critical band model is of fundamental importance in predicting the detectability of a signal as a function of the frequency spectrum of the signal and its masker. If the masker spectrum is outside the critical band centred on the signal, it does not have much effect on the detectability of the signal. As the centre frequency of the masker approaches that of the signal, detectability drops [1]. Detectability is poorest when the centre frequency of the masker is the same as that of the signal. An exception occurs when the signal and masker are both tones and are close enough in frequency to give

audible beats (within the same critical band - Version 2). Under these conditions, detection is poorest when the frequencies of the signal and masker are a few tens of hertz apart. When a masker is loud (more than 60 dB SPL (Sound Pressure Level)), its masking effect spreads upward in frequency. The louder it is, the more pronounced the upward spread of masking.

With a broad-band masker, detectability increases with increasing signal bandwidth for a signal of a given power per unit bandwidth, and as the number of components in the signal increases for a signal composed of sinusoidal components. The greatest improvement occurs when all the signal components are in the same critical band, because the auditory system integrates energy perfectly within a critical band. If the components are in separate critical bands, detectability also increases with the number of components, but the proportional increase in detectability is less. Detectability of such signals can be predicted best by assuming that the outputs of the stimulated critical bands are linearly combined [4]. For noise signals whose bandwidth exceeds a single critical band, detectability increases with increasing bandwidth at 1.5 dB/octave, a rate appropriate for a filter whose bandwidth is matched to that of the signal [5].

The auditory system seems to be more effective in detecting broad-band or multi-component signals than tones or narrow-band signals. In addition, since it appears able to adjust its listening bandwidth to match the spectrum of the incoming signal, it seems especially adapted for processing signals of unpredictable bandwidth.

The duration of the incoming signal is also an important determiner of its detectability. Ideally, the detectability of a signal should be a function of the total energy in the signal, which means that the detectable power should decrease by 3 dB for every doubling of signal duration. However, Green, Birdsall, and Tanner [6] found that this was true only for signal durations between 20 and approximately 200 milliseconds. For shorter tones, the specification of threshold as a function of duration is difficult because the signal energy is spread over a widening frequency band (frequency splatter).

For signal durations between 200 milliseconds and several seconds, the signal intensity required to maintain constant detectability decreases 1.5 dB per doubling of duration [6]. For longer signals, detectability is usually a function of signal power rather than signal energy. However, a frequent, but undocumented finding is that long duration signals with slow onset ramps are harder to detect than short bursts of the same power.

In sum, for effective detection, signals should be presented for at least 20 milliseconds, but presentation durations longer than 200 milliseconds do not improve performance. Time compression that results in fast onsets and offsets may improve the detection of signals that vary slowly in amplitude. If the time compression is of a kind that maintains the SNR, then as much of the signal as possible should be compressed into a 200 milliseconds duration.

A listener's knowledge about the stimuli can also affect a signal's detectability. In the laboratory, the listener usually knows the approximate onset time and has some knowledge about the frequency spectrum of the signal. The masker, if it is noise, varies randomly. If the listener can be provided with precise information about the signal or masker, the detectability of the signal may be raised as much as 10 dB [5,7]. In an operational setting, the operator knows very little about the characteristics of the target. Based on laboratory studies that randomly varied certain characteristics of the signal [8], Stallard and Leslie [9] predicted that the signal to noise ratio at which a signal could be detected would be 5.4 dB lower in the typical laboratory experiment than in an operational setting.

Not only does uncertainty about the signal reduce its detectability, it also reduces the probability that the operator will guess incorrectly that a signal was present (make a false alarm). Detection behaviour as a function of signal energy becomes much more threshold-like than when the signal is well specified to the observer. With uncertainty, the signal either is heard clearly or is not heard at all. Fluctuations in the observer's detection criterion have little effect, as compared to their importance under conditions of good information [10].

2.2. Simple Discrimination

Once a signal has been detected, it can be discriminated from another signal or from a remembered representation of a target signal. In the simplest cases, the discrimination is simply of amplitude or of frequency. Under ideal conditions, intensity shifts of 0.5 dB and frequency shifts of 0.1% can be discriminated. Both of these values are essentially independent of signal frequency for frequencies between 250 and 3000 Hz. Above 3 kHz, frequency discrimination deteriorates, but amplitude discrimination remains constant up to at least 8 kHz [11]. Near the threshold for detecting a change, the ear can tell there has been a change, but cannot tell in which direction it went. For reliable detection of the direction of a frequency change, the required frequency shift will be closer to 0.25% for frequencies below 4000 Hz, increasing to 7% at frequencies around 10,000 Hz [12].

As with detection, discrimination of a signal is affected by masking and by signal duration. Both amplitude and frequency discrimination improve as SNR increases from signal threshold intensity to 60 dB [13,14]. Most of the improvement is achieved in the first 15 dB above the signal threshold. Both amplitude and frequency discrimination are essentially independent of signal duration for signals longer than 200 milliseconds. Below approximately 200 milliseconds, discrimination decreases with decreasing duration [15,16].

For optimal discriminations of the frequency and amplitude of tones, the signals should be presented somewhere in the frequency range of 250 Hz to 3 kHz, at a high SNR, and for at least 200 milliseconds. In operational conditions, signal frequencies and durations can be transformed, but high SNR can rarely be achieved.

Discrimination has another meaning, distinct from detection of changes: it can mean the ability to discriminate one member from an ensemble. A complex tone consists of a series of harmonics. With practice, people can hear independently the lower harmonics of a complex tone [17]. These harmonics are separated by more than one critical band (Version 1). Tones that fall within a critical band ordinarily are heard as a single beating tone, but can be distinguished under several presentation conditions: (i) if the frequency of one tone varies slightly over time in a way different from the other(s), (ii) if the onset or offset times differ by more than about 30 milliseconds [18], or (iii) if one tone can form part of an ongoing sequence, such as a melody or in any other way can be connected differently with outside sounds [19]. Under these conditions the separated tonal complexes are perceived as coming from separate sources.

One essential element of the sonar operator's task is to segregate the sea sounds into separate sources. The TASS aids this task by separating sounds into their different directions. The ear is capable of coordinating many components into a single "voice," and of separating one voice out of many that may come from a given direction. Some of the important variables that contribute to this ability have been mentioned above, but the total range of important variables is unknown. Studies under simulated operational conditions are required to determine what the important variables might be. Bregman and his associates have done considerable work on the factors that lead a single tone to associate with specified other tones in an ensemble, but their results apply directly only to simple tonal complexes [20].

2.3. *Binaural Processing*

Humans are able to determine not only what sounds are in the environment, but from what direction they come. They do this through two completely independent mechanisms. One mechanism depends on comparison of timing and amplitude of sounds at the two ears, the other on reflections from the external ear (the pinna) that occur at each ear independently.

With headphones, internal reflections in the cavity formed by the headphone and the ear canal override any monaural pinna reflection cues to signal direction. In effect, sounds through headphones seem to come from an impossible direction in space. Only the interaural timing delay and amplitude differences can be used to suggest direction, but since the cues that normally occur for sounds in any external direction are absent, the brain resolves the dilemma by perceiving the sound as emanating from a location within the head, on an arc connecting the two ears. It is possible, by using excessive timing delays or unrealistic phase shifts such as phase inversions, to generate the illusion of a sound placed on the same arc but outside the ear rather than between the ears. The illusion of sound placement inside the head is probably augmented by the fact that relative phases and amplitudes do not change when the listener's head turns. If the sound came from outside the head, any change in head orientation would be accompanied by a corresponding change in interaural phase and amplitude difference.

The auditory system is as sensitive to amplitude differences between the ears as it is to amplitude changes in tones at one ear. Over the frequency range of 200 Hz to 10 kHz, amplitude differences of about 0.5 dB can be detected. Interaural time delays of as little as a few microseconds can be perceived by highly trained observers under the right conditions. In general, optimum discrimination occurs with signals of approximately one second in duration whose spectral content is primarily below 1500 Hz. For tones above 1500 Hz the discrimination of interaural time differences is impossible. With broad-band noises and complex sounds, temporal discrimination is a function of the amount of low frequency information available.

Differences in amplitude and interaural delay are perceived as a difference in position of the sound "image" in the head. If precisely the same information is presented to the two ears, a single fused image is heard in the centre of the head. Changing the amplitude or the relative onset time of the sound at one ear causes the position of the image to shift (referred to as lateralization of the image). With sufficient disparity between the two ears the image becomes diffuse and in some cases, splits into two. Through analysis of direct and indirect data on lateralization, it is possible roughly to correlate image position with amplitude ratio and temporal delay [21].

The effect on image position of changing interaural amplitude is generally straightforward. Similar effects are found whether the sound is a high frequency tone, a click, a noise pulse or a pulse train. As the intensity ratio increases from zero to approximately ten dB, the image is perceived as shifting from the centre of the head to the ear with the more intense sound. At least for a noise burst, the results are independent of the overall level and duration of the stimulus [22].

The effects of interaural time delays are much more complex. For tonal signals, the effect of changing interaural delay depends on the frequency of the signal. Above 1500 Hz, the position of the image is not affected by altering the interaural time difference. Below 1500 Hz, interaural phase difference is important, rather than interaural time difference. When the signals in the two ears are in phase, there is a clearly fused central sound image. As the phase difference increases, the fused image moves toward the leading ear, but as the phase difference approaches 180°, the image becomes more and more diffuse and difficult to locate. As the phase shift increases, the stimulus is equivalent to a lead at the other ear, and the image fuses again at the other side of the head [23]. The amount by which the fused image moves from the centre of the head is a function of frequency: near 1500 Hz there is little movement, no matter what the phase, but at 700 Hz or below, the image can move as far as the ears, or even beyond if the phase shift corresponds to an "unnatural" time delay.

For a continuous broad-band noise, a click, or similar type of waveform, lateralization depends on interaural time delay rather than phase. As the interaural delay increases from zero to one millisecond, the image shifts from the centre of the head to the leading ear. For continuous broad-band noises, as the interaural delay increases beyond one millisecond, the image remains at the leading ear but becomes more diffuse until it fills the

whole head [24]. At this point, it is indistinguishable from the image produced by statistically independent noises in the two ears. Increasing interaural delay beyond one millisecond with click stimuli causes the image to split into two [25]. With an interaural delay of about 15 milliseconds, the images are located at the two ears. The figure of one millisecond is associated with the headwidth delay while the figure of 15 milliseconds is thought to reflect interaural "memory" [21].

Natural sounds differ in both interaural timing and interaural amplitude, as well as having monaural pinna reflections appropriate to their directions. The absence of the monaural pinna reflections gives a perceived localization within the head, but the interaural amplitude and timing differences must still be mutually consistent if the sound image is to have a unique and clear location. If the interaural amplitude difference is inconsistent with the interaural timing difference for a particular sound, the perceived location will be diffuse, or may even be in the form of a double image. Processing of sea sounds from a given beam must ensure that the binaural presentation has compatible values for interaural amplitude and timing differences.

2.4. Change Rates of Amplitude and Frequency

Most of the studies discussed above measure detection and discrimination of independent sounds that differ along one or more dimensions. Frequently, the problem is to detect or discriminate a change in frequency or amplitude of a single sound over time. The ear cannot detect very slow glides (smooth frequency or amplitude changes), and very fast glides cause a momentary "frequency splatter," that sounds like a click rather than like a glide. Between these two extremes there is a range of highly detectable glide timings that the ear can use for discrimination, and that are highly detectable.

Speech is the type of complex sound for which the ear seems best attuned. In speech, much information is carried by changes that occur over a time range of between about 20 and 70 milliseconds. Faster changes are perceived as "instantaneous", and slower changes are not significant.

For signals simpler than speech, several experiments have investigated the detection of frequency and amplitude variation as a function of rate, and the discrimination of change rate itself has also been studied. Typical experiments do not cover the slow changes likely in sea sounds; experiments tend to cover modulation rates of 1 Hz or faster. These studies generally suggest that the detection of modulation (either of frequency or of amplitude) is unaffected by modulation rate up to a rate of around 20 Hz, but is poorer for higher modulation frequencies. Hence, one should expect no decline in detection or discrimination performance for time compression of repetitive signal changes (e.g. propeller blade noise) so long as the transformed modulation frequency remains below about 20 Hz.

The discrimination of glide rate (which of two glides is faster) has been little studied. Nabalek and Hirsh [26] investigated tones of low (250 Hz), medium (1 kHz) and high (4 kHz) frequency. For most conditions tested, discrimination would be best when the glide rate was approximately $0.01 f_0$ Hz/millisecond over a 30 millisecond period (where f_0 is the base frequency of the glide).

3. PROCESSING OF SONAR SOUNDS FOR AURAL DISPLAY

Anecdotal evidence suggests that trained operators can sometimes make quite accurate target identifications simply by listening to unaltered sounds from the sea. However, extensive training is usually required before an operator can readily recognize these unaltered sounds. Most of the sounds radiating through the ocean are low in frequency with broad-band, narrow-band and time-varying components [27].

The ear deals best with sounds between about 300 Hz and 3000 Hz. In this range, detection thresholds are lowest, amplitude and frequency discrimination is finest, and lateralization best. It is good at detecting small changes in frequency and amplitude of short duration signals. Only occasionally does extending presentations beyond 200 milliseconds provide any benefit to auditory processing (binaural fusion is an exception, for which a limit of about one second is probably suitable). Whereas the ear is limited as to how narrow a band of frequencies it can monitor, it is able to adjust the bandwidth of its internal filters to optimize the signal to noise ratio for multi-component signals and broad-band signals. The ear is exquisitely sensitive to complexes of components that have common onsets and offsets, or that vary coherently in frequency. Such complexes can be extracted from backgrounds consisting of similar complexes that differ in timing or frequency variation, and identified almost as if the background complexes did not exist. The ear is not good at distinguishing complexes that have synchronous timings or common patterns of frequency variation.

Processing should aim at transforming slowly changing and low frequency sounds into the timing and frequency range for which the ear is optimized. We know of no work so far that has attempted to transform sea sounds into different timing and frequency ranges. Proper processing should produce sounds that the operator can discriminate more easily and may add to the information available in the raw sound.

A "good" transformation should result in a signal that is statistically much like speech in both its spectral contour and the rate at which its strong frequencies vary. Such a transformation might be fixed over time, based on prior knowledge of the characteristics of interesting targets, or it might be adaptive, based on the recent history of its output. One component would almost certainly be a large time compression with no corresponding frequency shift (since the major target components detected by a TASS match the ear's capabilities quite well). Another component would necessarily be spectral shaping to compensate for the self-noise of the tow (adjusted for the ambient

noise in the operator's working area).

Time compression can be performed without frequency shifting, by operations on a time-frequency analysis of the signals. Any waveform analyzed by a bank of overlapping filters can be resynthesized exactly from the filter output energy functions, disregarding their phases. Time compression can therefore be achieved by low-pass filtering, or by selectively sampling the energy output waveforms of a bank of filters, and then using a time-compressed version of those output waveforms to control a bank of oscillators having frequencies corresponding to the individual filters. The resulting audio signal should have a periodicity pitch that is raised by a factor equal to the time compression, but a spectrum-height "pitch" unchanged from the original.

If it is desired that the periodicity pitch be unchanged through the time compression, this can be done by an extension of the same scheme, although the processing is more complex.

The actual value of preprocessing along the lines discussed above remains to be tested. Although the value of shifting signals is supported by basic research it has not been tested out to any extent. Only recently has it become feasible to manipulate spectral and temporal components of signals independently.

4. INTEGRATION OF INFORMATION

We turn now from consideration of auditory displays alone to the integration of auditory with visual displays. Because of the high information rates from a TASS, displays that improve performance at the cost of time are impractical. An aural display will be effective to the extent that it enhances the processing possible for the operator in the time he would ordinarily devote to the visual display. Information from appropriately related sources can be effectively combined, whether in one modality or two, if the information is perceived as being useful. Detection and classification can be enhanced by providing complementary information to the two senses, or through integration of the output of automatic processing with sensory processing. The human can also effectively utilize cues and templates that provide examples of known targets or features of targets.

4.1. Bisensory Displays

The concurrent visual and aural presentation of complex signals was studied by Halpern [28]. He concentrated on determining some of the conditions in which a bisensory presentation would be superior to a unisensory presentation of the same information in an identification task. He also varied the type of information presented visually and the duration of the stimulus. Performance in the bisensory conditions was consistently superior to the unimodal conditions. Using a time-history visual display and an unprocessed aural presentation of the signal, subjects identified 75% of the targets with the

visual display, 85% with the aural display and over 95% with the sensory display.

In general, Halpern [28] found that bisensory presentation was best when different information was presented to each sense and when the user had time (in his study, 5 seconds) to integrate the information from both displays.

The value of bisensory displays for detection was examined by Lovelless, Brebner, and Hamilton [29]. They asked subjects to detect a signal masked by noise, when presented aurally, visually (on an oscilloscope) or both. The results showed that a bisensory presentation could enhance the detectability of signals. The benefits were not as substantial as those found by Halpern [28], but in their study, the information was the same in both displays.

4.2. Automatic Processing

A different form of integration involves combining the output of a display system with the result of computer processing of the same information. Many sonar systems incorporate an automatic alert system that analyses the incoming sounds, and alerts the operator either visually or aurally (or both) when the probability of a threat target exceeds some threshold. The usefulness of such a system depends in part on the effectiveness of the analysis algorithms and the extent to which the operator effectively integrates the computer processing with his own analysis. The development of effective algorithms is not within the scope of this paper, although a knowledge of how the human processes complex sounds might result in the development of more effective algorithms. Studies on computer-assisted detection and identification indicate that an operator can use the information provided by a computer, but not optimally.

Ingleby, Hamilton, and Copeman [30] and Hartley, Olsson, and Ingleby [31] examined the effect on performance in a vigilance detection task of providing an automatic alert that was correct 75% of the time. In general, the provision of an alert resulted in an increase in the number of correct detections but did not result in improved discrimination of signal from noise. The increase in correct detections was accompanied by a corresponding increase in false alarms. In addition, subjects were more willing to accept the computer's decision when it indicated signal present than when it indicated signal absent. With practice, the subjects placed increasing reliance on the computer decisions.

Corcoran, Dennett, and Carpenter [32,33] studied the effect of providing computer assistance on classification performance with an aural display. The computer display showed the probability that the sound had been generated by each of the potential sources. In each session, performance was measured without the computer analysis, then with it, and then again without it. The variables were signal to noise ratio and the reliability of the computer analysis. In all conditions, subjects' performance with the computer analysis was equal or superior to their performance without it. The computer analysis

was most helpful when the probability of the computer identifying the target correctly and the probability of the listener identifying the target correctly without the computer were both about 50%. When the probability of the computer correctly identifying the signal was close to 100%, the subjects' performance in the trials without the computer did not improve across sessions.

Operators can learn to integrate evidence from the computer with the evidence from their senses. However, optimum performance is not easily achieved. If the computer is seen as accurate, the aural input is likely to be ignored; if the computer is seen as less reliable than the operators' own senses, its evidence is likely to be ignored.

4.3. *Cueing*

When an operator analyses a sound, he presumably compares it with stored representations of previously identified sounds. The better the set of stored representations the more accurate will be his classification of the new sound. Classification might be aided by providing labelled templates or cues for the operator to use as comparisons. These cues might be written descriptions, or aural or visual templates.

Most of the applied studies on cueing have been training studies. For instance, Corcoran, Carpenter, Webster, and Woodhead [34] examined the effect of different forms of verbal description on learning to identify simulated sonar targets. They found that providing listeners with an acoustical description of each target resulted in superior and more rapid learning than providing listener-generated descriptions, or one word labels. Other studies [35,36] found acoustical description of little value. The reason for the difference is not clear but it is possible that differences in the stimuli in the different studies may have contributed. The descriptions provided by Annett [35] provided more information than the listeners needed while the stimuli of Woodhead, Carpenter, and Webster [36] were relatively simple.

The value of providing aural templates to experienced operators was reported by Mackie [37]. Operators were provided with a tape of signal samples to use when classifying unknown targets. Classification performance was compared with and without this memory aid. The operators did significantly better with the aid, the greatest improvement being found with those operators who did poorest initially.

The effective use of cueing probably requires considerable thought. As with other sources of information, cues are useful to the extent they provide information not otherwise available to the operator. Since templates require additional processing time, it may be cost effective only for inexperienced operators. It is proposed below that templates be made available for alternation with sea sounds at the discretion of the operator.

5. PROPOSED AURAL DISPLAYS

Multiple sources of information seem most useful when the different sources provide complementary and believable information. The ear has been shown to handle multi-component, band-limited and rapidly varying sounds best. Visual displays, in contrast, have proven best at presenting narrow-band steady-state sounds. Sounds found in the ocean do contain both broad-band and time-varying components [27], but these have largely been ignored because of the use of narrow-band filters and long processing times with visual displays. The diagnostic properties of broad-band and time-varying sounds for target identification are not known, but it would be surprising if they did not add evidence to the information available from the narrow-band signals shown on the visual displays. In the past, extensive training has been required to utilize the potential information in aural displays. Since most of the sounds radiating through the ocean are low frequency and vary slowly over time, the operator had to learn to discriminate signals whose characteristics differ from the kinds of sounds the ear processes most effectively. Through appropriate transformations, the sea sounds can be shifted into the timing and frequency ranges for which the ear is optimized. Thus, it should be possible to develop aural displays that can be used effectively and efficiently with visual displays for both detection and identification.

5.1. Identification

For identification, an aural display should be used to provide information complementary to the narrow-band steady-state information presented on the FTI display. It would be used for analysing broad-band and time-varying components of the incoming sound.

For optimum perception of the time-varying components, the incoming information should be time compressed. On the assumption that changes in sea sounds occur on a time-scale of seconds or minutes, time compressions of one to three orders of magnitude should be useful. A factor of 100 might be normal. Properly done, such compression might improve SNR by 10 or 20 dB as well as improving the timing characteristics of the signal for the human ear. A time compression of such magnitude should permit the operator to review the history of the signal source or to compare it with a library of aural templates. While provision of templates may improve the effectiveness of an operator, it will reduce his efficiency unless the comparison time is kept to a minimum. Severe time compression achieves this end.

Since it is undesirable to present a sound for much longer than one second, something has to change in the display quite regularly. It is useful to think of a time-compressed display in a template-matching mode; in every other second, a template of a suspected target is presented in alternation with the sound from one or more beams. The intensities of lines in the template should, of course, be adjusted to be close to the corresponding intensities in the real sea sound. If a target matching the template is present, it should then be perceived as standing out from the complex of noises in the

sea. If, on the other hand, a false target similar to the suspected one is present, one or more components should appear to drop out or to shift in frequency during the alternation.

By manipulation of relative interaural time delays and intensity differences, sound from any one beam could be placed anywhere in the head. Three or more beams could simultaneously be located in distinct positions. Beam patterns overlap in the sea, and with appropriate selection the beams in the head could be given a similar degree of positional overlap. In such a configuration, targets crossing beam boundaries would not be perceived as doing anything awkward, as happens on a visual display. They would move smoothly through the auditory space, rather than appearing on separate panels of a multi-panel visual display.

Such a presentation could give a finer estimate of bearing than is easy to get with a visual display, by virtue of the fact that a target off-center for any beam will appear to some degree in the neighbouring beam. Thus it will be localized in aural space between the nominal locations of the corresponding beam centres.

5.2. *Detection*

Systems for initial detection include automatic alert, BTI displays, and multiple FTI displays. The automatic processing and the FTI displays provide information on targets radiating stable narrow-frequency sounds whereas the BTI can provide broad-band information. The varieties of information provided by an aural display are not available, except perhaps for the broad-band information provided by the BTI. The BTI information is undifferentiated; if there is energy from the particular bearing, it will be displayed regardless of its frequency, bandwidth, or stability. The ear can adjust the bandwidth of its listening filters to match the bandwidth of the incoming sounds. It can also separate out different sources of information. Both the aural display and the BTI can be used to give a directional history of signal sources.

For detection, a display must allow efficient monitoring of all the incoming channels. A time-compressed display should permit the operator to scan the whole ocean in real time. With the compression ratios suggested in this report and by taking advantage of the ear's localization ability, the scan might well include all beams and leave time for other functions as well. During an ocean scan, the locations of the beams could be moved from left to right across the head, new beams being introduced at the left and old ones departing at the right. It is likely that the direction of scan would have to be reversed frequently, to avoid auditory motion after-effects or even the possibility of operator dizziness. Since the beams cover only an ambiguous 180° sector of the ocean rather than 360°, a suitable scan pattern might move from beam 1 to beam N and then back, rather than continuing in one direction.

A scanning aural display might be particularly effective as an adjunct to an automatic alert system. The automatic alert would cue the operator to pay attention to specific channels. The operator would compare the output of these channels with adjacent non-cued channels. If the operator impression based on the aural display supported that of the automatic alert then the channel could be marked for a more detailed look, using the visual displays or further enhancements of the aural display.

6. CONCLUSIONS

The auditory system is well suited for analyzing time-varying and broad-band sounds, but the time-scale of variation must be on the order of tens of milliseconds to be useful. The time variation of sea sounds has not been much used for detection or identification of targets, most effort having historically been concentrated on the use of stable lines with visual displays. Compressing the incoming signal in time would provide a signal well matched to the auditory system. Such transformations would allow presentation of a greater range of information than is possible with traditional aural or visual displays. The additional information could be other channels in the towed array system or stored samples of targets. The human can effectively integrate multiple sources of information under the right conditions. These sources may be sensory information that must be coordinated and attributed to entities in the outer world, or they may be computer-based suggestions as to interpretations of incoming information. The addition of aural displays optimized to match the capabilities of the human should lead to effective and efficient processing of the information provided by towed array systems.

7. REFERENCES

1. PATTERSON, R. D., and D. M. GREEN. Auditory Masking. In E. C. Carterette, and M. P. Friedman (Eds.), *Handbook of Perception Volume IV: Hearing*. New York: Academic Press Inc., 1978.
2. TAYLOR, M. M., S. M. SMITH, and M. E. TYTLA. Auditory Filter Shape by Subthreshold Summation. *J. Acoust. Soc. Amer.*, 57, S78(A), 1975.
3. FETH, L. L., and H. O'MALLEY. Two-Tone Auditory Spectral Resolution. *J. Acoust. Soc. Amer.*, 62, 940-947, 1977.
4. GREEN, D. M. Detection of Multiple Component Signals in Noise. *J. Acoust. Soc. Am.*, 30, 904-911, 1958.
5. TAYLOR, M. M., and S. M. FORBES. Monaural Detection with Contralateral Cue (MDCC). I. Better than Energy Detector Performance by Human Observers. *J. Acoust. Soc. Amer.*, 46, 1519-1526, 1969.
6. GREEN, D. M., T. G. BIRDSALL, and W. P. TANNER. Signal Detection as a Function of Signal Intensity and Duration. *J. Acoust. Soc. Am.*, 29, 523-531, 1957.
7. SPIEGEL, M. F., and D. M. GREEN. Signal and Masker Uncertainty with Noise Maskers of Varying Duration, Bandwidth, and Center Frequency. *J. Acoust. Soc. Amer.*, 71, 1204-1210, 1982.
8. GREEN, D. M. Detection of Auditory Sinusoids of Uncertain Frequency. *J. Acoust. Soc. Am.*, 33, 897-903, 1961.
9. STALLARD, J. M., and J. M. LESLIE. *Psychoacoustics and Passive Sonar Detection*. NOLTR 74-27, Naval Ordnance Laboratory, Silver Spring, Maryland 20910, 1974.
10. TAYLOR, M. M. Detectability Theory and The Interpretation of Vigilance Data. *Acta Psychologica*, 27, 390-399, 1967.
11. JESTEADT, W., C. C. WIER, and D. M. GREEN. Intensity Discrimination as a Function of Frequency and Intensity Level. *J. Acoust. Soc. Amer.*, 61, 169-184, 1977.
12. HENNING, G. B. Frequency Discrimination of Random-Amplitude Tones. *J. Acoust. Soc. Amer.*, 39, 236-339, 1966.
13. HENNING, G. B. Frequency Discrimination in Noise. *J. Acoust. Soc. Am.*, 41, 774-777, 1967.
14. HENNING, G. B., and S. L. BLEIWEAS. Amplitude Discrimination in Noise. *J. Acoust. Soc. Am.*, 41, 1365-1366(L), 1967.

15. HENNING, G. B. A Comparison of the Effects of Signal Duration on Frequency and Amplitude Discrimination. In R. Plomp, and G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing*. Proceedings of the International Symposium on Frequency Analysis and Periodicity Detection in Hearing, Driebergen, The Netherlands, June 23-27, 1969, Leiden: A. W. Sijthoff, 1970.
16. HENNING, G. B., and J. PSOTKA. Effect of Duration on Amplitude Discrimination in Noise. *J. Acoust. Soc. Am.*, 45, 1008-1013, 1969.
17. PLOMP, R. The Ear as a Frequency Analyzer. *J. Acoust. Soc. Am.*, 36, 1628-1636, 1964.
18. RASCH, R. A. The Perception of Simultaneous Notes such as in Polyphonic Music. *Acustica*, 40, 21-33, 1978.
19. STEIGER, H., and A. S. BREGMAN. Competition Among Auditory Streaming, Dichotic Fusion, and Diotic Fusion. *Perception and Psychophysics*, 32, 153-162, 1982.
20. BREGMAN, A. S. The Formation of Auditory Streams. Chapter 4 in J. Requin (Ed.), *Attention and Performance VII*. Proceedings of the Seventh International Symposium on Attention and Performance, Senanque, France, August 1-8, 1978, Hillsdale, N. J.: Lawrence Erlbaum Associates, 1978.
21. DURLACH, N. I., and H. S. COLBURN. Binaural Phenomena. Chapter 10 in E. C. Carterette, and M. P. Friedman (Eds.), *Handbook of Perception Volume IV: Hearing*. New York: Academic Press Inc., 1978.
22. PINHEIRO, M. L., and H. TOBIN. Interaural Intensity Difference for Intracranial Lateralization. *J. Acoust. Soc. Amer.*, 46, 1482-1487, 1969.
23. SAYERS, B. Acoustic-Image Lateralization Judgments with Binaural Tones. *J. Acoust. Soc. Amer.*, 36, 923-928, 1964.
24. BLODGETT, H. C., W. A. WILBANKS, and L. A. JEFFRESS. Effect of Large Interaural Time Differences upon the Judgement of Sidedness. *J. Acoust. Soc. Amer.*, 28, 839-843, 1956.
25. BABCOCK, H., and S. SUTTON. End Point of Lateralization for Dichotic Clicks. *J. Acoust. Soc. Amer.*, 39, 87-102, 1966.
26. NABELEK, I., and I. J. HIRSH. On the Discrimination of Frequency Transitions. *J. Acoust. Soc. Amer.*, 45, 1510-1519, 1969.
27. URICK, R. J. *Principles of Underwater Sound*. New York: McGraw-Hill Inc., 1975.
28. HALPERN, J. *Bisensory Information Processing: Discrimination as a Function of Cue Redundancy and Task Difficulty*. TR-6, Computer Image

Corporation and University of Denver, Department of Psychology, 1971.

29. LOVELESS, N. E., J. BREBNER, and P. HAMILTON. Bisensory Presentation of Information. *Psych. Bull.*, 73, 181-199, 1970.
30. INGLEBY, J. D., P. HAMILTON, and A. K. COPEMAN. *Simulated Computer Assistance in a Signal-Detection Task (Its Effectiveness as a Function of Signal Probabilities, Cueing Time and Knowledge of Results)*. OES-4/72, Royal Naval Personnel Research Committee, London, England, 1971.
31. HARTLEY, L. R., R. G. OLSSON, and J. D. INGLEBY. *An Unbiased Assessment of Simulated Computer Assistance in a Signal Detection Task*. OES-1/74, Royal Naval Personnel Research Committee, London, England, 1972.
32. CORCORAN, D. W. J., J. L. DENNETT, and A. CARPENTER. Cooperation of Listener and Computer in a Recognition Task: Effects of Listener Reliability. *J. Acoust. Soc. Amer.*, 47, 480-488, 1970.
33. CORCORAN, D. W. J., J. L. DENNETT, and A. CARPENTER. Cooperation of Listener and Computer in a Recognition Task. II. Effects of Computer Reliability and "Dependent" versus "Independent" Conditions. *J. Acoust. Soc. Amer.*, 52, 1813-1819, 1972.
34. CORCORAN, D. W. J., A. CARPENTER, J. C. WEBSTER, and M. M. WOODHEAD. Comparisons of Training Techniques for Complex Sound Identification. *J. Acoust. Soc. Amer.*, 44, 157-187, 1968.
35. ANNETT, J. *Sonar Recognition Training: An Investigation of Whole VS Part and Analytic VS Synthetic Procedures*. NAVTRADEVEN 87-C-0105-1, Naval Training Device Center, Orlando Florida, 1971.
36. WOODHEAD, M. M., A. CARPENTER, and J. C. WEBSTER. *Learning to Identify Sounds Using Explicit Instructions*. RNPRC OES-74/5, Royal Naval Personnel Research Committee, London, England, 1973.
37. MACKIE, R. R. *Research on Factors Influencing The Interpretation Of Sonar Signals*. 778-6, Human Factors Research, Incorporated, 6780 Cortona Drive Goleta, California 93017, 1974.

POOR

QUALITY

ORIGINAL